

A Simple Unsupervised Color Image Segmentation Method based on MRF-MAP

Qiyang Zhao

Abstract

Color image segmentation is an important topic in the image processing field. MRF-MAP is often adopted in the unsupervised segmentation methods, but their performance are far behind recent interactive segmentation tools supervised by user inputs. Furthermore, the existing related unsupervised methods also suffer from the low efficiency, and high risk of being trapped in the local optima, because MRF-MAP is currently solved by iterative frameworks with inaccurate initial color distribution models. To address these problems, the letter designs an efficient method to calculate the energy functions approximately in the non-iteration style, and proposes a new binary segmentation algorithm based on the slightly tuned Lanczos eigensolver. The experiments demonstrate that the new algorithm achieves competitive performance compared with two state-of-art segmentation methods.

Index Terms

Image segmentation, Markov random fields, maximum a posteriori, unsupervised segmentation.

I. INTRODUCTION

Unsupervised color image segmentation is important in various image processing and computer vision applications, such as medical imaging [1], image retrieval [2], image editing [3], and object recognition [4]. Estimating the maximum a posteriori (MAP) on the Markov random fields (MRF), is so far an fundamental tool which is widely adopted both in the unsupervised color image segmentations and supervised ones [5]-[10]. In all existing MRF-MAP-based image segmentation methods, their goals are to find the optimal label configurations on pixels to maximize the posterior probability which is in

Qiyang Zhao is with the School of Computer Science and Engineering, Beihang University, Beijing, 100191, China. E-mail: zhaoqy@buaa.edu.cn.

proportion to the product of the MRF priors and the likelihoods terms, or equivalently, to minimize the energy function of the smoothness terms plus data terms.

The likelihood/data terms consist of the parameters of some color distribution models, here these models specify the probabilities of any color occurring in each segmentation. These parameters usually derive from user interactions or random sampling, hereby there is always non-negligible inaccuracy in the likelihood terms. In order to address this, the Expectation-Maximization (EM), simulated annealing and other iterative methods are usually adopted to progressively approach the appropriate parameters, especially in the case of unsupervised segmentations up to today [5]-[7]. There are many choices of optimization algorithms to be adopted in the M-step [11].

There are three major disadvantages in the current unsupervised segmentation methods. The first is the low efficiency of the iterative frameworks, particularly when faced with large size images. The second is always on the high risk to be trapped to the local optima of the energy function. Although stepping out and restarting the iteration is a reasonable improvement, there would inevitably be additional computational load and it is possible to be trapped again. The third is the coarseness in the segmentation results, and it is partially caused by the roughness of the likelihood parameters. To address these issues, the letter proposes a new unsupervised binary segmentation method based on the approximation of the likelihood terms, where the iterative computations are replaced by only one single step of solving the eigenvector of the largest eigenvalue, therefore the computational efficiency is remarkably improved. This new approach increases the chance to high quality segmentation results by obtaining the nearly optimal solutions to maximize the posterior probability as possible. It also provides us an effective way to test and verify the MRF prior parameters or their involved generating schemes, which are also critical to the segmentation tasks.

II. COLOR IMAGE SEGMENTATION BASED ON MRF-MAP

In the following sections, we focus on the binary segmentation with the label set $\{fore, back\}$. The computational goal of these methods is to maximize the probability $P(L|I)$ of the segmentation label configuration L given an image I . According to the Bayesian rule, it is equivalent to maximize the joint probability $P(I, L) = P(I|L) \cdot P(L)$, where the prior $P(L)$ is established on the Markov random field of I , the conditional probability $P(I|L)$ is the likelihood that the pixel colors occur in their corresponding segments marked by different labels. In a more prevailing view, what we need is to minimize a energy function E which is the negative log-hood of $P(I, L)$. Here E is usually written in the form of a data

term E_D plus a smoothness term E_S multiplied by a factor λ :

$$E = E_D + \lambda E_S \quad (1)$$

where E_D reflects the likelihood of the color occurrences in the image segments, E_S is the sum of all adjacency interaction potentials of each two neighboring pixels of different labels:

$$E_D = \sum_p -\ln P_{L(p)}(p), E_S = \sum_{\substack{(p,q) \in N \\ L(p) \neq L(q)}} S(c(p), c(q)) \quad (2)$$

where $L(\cdot)$ is the pixel label, $c(\cdot)$ is the pixel color, and $S(\cdot, \cdot)$ is the perceptually similarity weight of two colors. It is meant two pixels p and q are adjacent to each other by noting $(p, q) \in N$. Thereafter, the segmentation task is to pursue an appropriate a label configuration to reach the lowest energy. Although as indispensable as the data terms when computing the energy functions, the smoothness terms are not to be addressed in the letter.

There are two steps when determining the coefficients in the data term E_D . First, a suitable color statistical model should be chosen. Histograms are usually adopted for images of small color spaces, such as gray scale or 256 colors, but it is not suitable for large color spaces as the samples were statistically too few when facing so many histogram bins. Some other models fit large color spaces well, and make an appropriate compromise between the efficiency and accuracy, such as the Gaussian mixture model (GMM). Second, the model parameters should be determined. However here arises a *chicken or the egg* dilemma unavoidably: we have to know the parameters first to minimize the energy function to obtain the optimal segmentation, but the optimal segmentation is just the key to produce the accurate parameters mentioned above. The usually adopted solutions to this are the iterated procedures, such as EM, in which the estimation and optimization are performed sequentially but isolatedly in each single loop. Here the initial parameters are determined from sample pixels chosen by user interactions or random samplings. There are many choices to perform the optimization [11]: graph cut, Loopy Belief Propagation (LBP) and Iterated Conditional Model (ICM).

The low computational efficiency is an adherent shortcoming of the iterated solutions, as it is extremely hard to predict when and where the iterations would halt. Furthermore, although the minimum of the energy function is an unambiguous target itself, the actual aim of the iterated solutions is not mathematically explicit for us to approach. As a result, the iterations are likely finished at the local minimal in most cases. To address these issues, the letter proposes an approximating expression which is rather close to E_D in (2), and associate the approximated target energy function with the cut on a

complete graph G which has both positive and negative edge weights. In the following manipulations, it is rather straightforward to solve an eigen-system to pursue the minimum cut C on G , so to minimize the energy. There the expected segmentations are worked out directly without considering the troublesome parameters of the data terms at all.

III. SEGMENTATION BASED ON APPROXIMATE MRF-MAP

Consider there are n pixels of $m(m \ll n)$ colors in the image I , and any two colors are perceptually distinguishable from each other. It is almost always achievable with the help of the existing color clustering algorithms, even facing much splendid images. Therefore we choose the histograms as our color distribution models in (2). Let $c(p)$ be the color of the pixel p . For all pixels having color i , let n_i be the total amount, $n_{f,i}$ be the amount of those having label *fore*, so be $n_{b,i}$ and label *back*. Clearly $n_i = n_{f,i} + n_{b,i}$. Let F and B be the two sets of *fore* and *back* pixels respectively, and the corresponding pixel amounts are $n_f = n_{f,1} + n_{f,2} + \dots + n_{f,m}$ and $n_b = n_{b,1} + n_{b,2} + \dots + n_{b,m}$. Leaving the smoothness term E_S unchanged, we have the the data term E_D as

$$\begin{aligned}
 E_D &= \sum_{p \in F} -\ln \frac{n_{f,c(p)}}{n_f} + \sum_{p \in B} -\ln \frac{n_{b,c(p)}}{n_b} \\
 &= (n_f \ln n_f + n_b \ln n_b) - \sum_{i=1}^m (n_{f,i} \ln n_{f,i} + n_{b,i} \ln n_{b,i}) \\
 &= [n \cdot (\frac{n_f}{n} \ln \frac{n_f}{n} + \frac{n_b}{n} \ln \frac{n_b}{n}) + n \ln n] \\
 &\quad - \sum_{i=1}^m [n_i \cdot (\frac{n_{f,i}}{n_i} \ln \frac{n_{f,i}}{n_i} + \frac{n_{b,i}}{n_i} \ln \frac{n_{b,i}}{n_i}) + n_i \ln n_i]
 \end{aligned} \tag{3}$$

Now consider a function $g(x)$ defined on the interval $[0, 1]$:

$$g(x) = \begin{cases} x \ln x + (1-x) \ln(1-x), & \text{if } 0 < x < 1. \\ 0, & \text{otherwise.} \end{cases} \tag{4}$$

Clearly g is continuous on the whole interval $[0, 1]$. Expanding the log terms into the Taylor series and simplifying the expression, we have $g(x) = -\frac{5}{2}xy + \Delta(x)$, where $y = (1-x)$ and

$$\Delta(x) = -xy \left[\frac{1}{3}(x^2 + y^2) + \frac{1}{4}(x^3 + y^3) + \dots \right], \tag{5}$$

which has the mean value $\int_0^1 \Delta(x) dx = -\frac{1}{12}$ on $[0, 1]$. If we replace $\Delta(x)$ with the constant $-\frac{1}{12}$ on the entire interval, the mean squared error would be $\int_0^1 (\Delta(x) - \frac{1}{12})^2 dx < 3 \times 10^{-4}$. Since the MSE is considerable small, it is totally acceptable to approximate g with $g^* = -\frac{5}{2}xy - \frac{1}{12}$, as shown in Fig. 1.

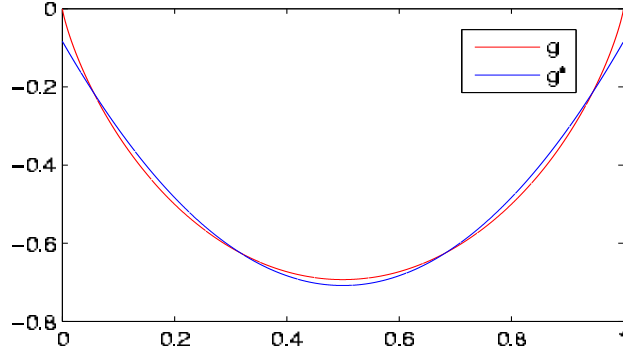


Fig. 1. Comparison of function g and its approximation g^* on $[0, 1]$.

With the help of g^* , now we can approximate the energy function E in (3) with

$$\begin{aligned} & \left(-\frac{5}{2n}n_f \cdot n_b - \frac{1}{12} + n \ln n\right) - \sum_{i=1}^m \left(-\frac{5}{2n_i}n_{f,i} \cdot n_{b,i} - \frac{1}{12} + n_i \ln n_i\right) \\ & + \lambda \cdot \sum_{\substack{(p,q) \in N \\ L(p) \neq L(q)}} S(c(p), c(q)) \end{aligned} \quad (6)$$

Since m , n and all n_i 's are constant in the input image, it is actually to minimize the following function E^* when minimizing the above expression as the approximation of E :

$$E^* = -\frac{5}{2n}n_f \cdot n_b + \sum_{i=1}^m \left(\frac{5}{2n_i}n_{f,i} \cdot n_{b,i}\right) + \lambda \cdot \sum_{\substack{(p,q) \in N \\ L(p) \neq L(q)}} S(c(p), c(q)). \quad (7)$$

Now construct an undirected complete graph G of n nodes each of which corresponds to one pixel, and set the edge weight $w(p, q)$ to be the sum of the following three terms

$$\begin{aligned} w_1(p, q) &= -\frac{5}{2n} \\ w_2(p, q) &= \begin{cases} \frac{5}{2n_i}, & \text{if } c(p) = c(q) = i, \\ 0, & \text{otherwise.} \end{cases} \\ w_3(p, q) &= \begin{cases} \lambda \cdot S(c(p), c(q)), & \text{if } (p, q) \in N, \\ 0, & \text{otherwise.} \end{cases} \end{aligned} \quad (8)$$

It is easy to prove that, for any label configuration L of the binary segmentations, E^* is equal to the capacity of the cut $C = \{F, B\}$ arising from L on G . Therefore it is equivalent to find the minimum cut on G when minimizing E^* . However the existing minimum cut algorithms are not suitable here because of the existence of negative edge weights. In fact, the problem here is computationally equivalent to a

well-known NP-complete problem, Max-Cut, on the graphs with non-negative weights. Therefore it is hard to obtain the exact minimum of E^* in polynomial time.

Our solution to this is to generalize it into the continuous real space R^n . First we put the label configuration into an indicator vector $D = [d_1, d_2, \dots, d_n]^T$: for $i = 1, 2, \dots, n$, let $d_i = +1$ if the i th pixel label is *fore*, and -1 for *back*. Then establish a matrix $W = [w(p, q)]$ and let S_W denote the sum of all its entries. It is easy to prove that the cut value is equal to $\frac{1}{2}(S_W - D^T W D)$. After generalizing the d_i 's to be in the continuous interval $[-1, 1]$ instead of $\{+1, -1\}$, our task becomes into

$$\max D^T W D, \text{ s.t. } \|D\|_2 = n \quad (9)$$

because S_W is a constant here. According to the *Lagrange multiplier* method [12], the solution to (9) is the eigenvector D^* corresponding to the largest eigenvalue of W . The *Lanczos* algorithm, well known as the fastest method solving extremal eigenvectors for large sparse matrices, is adopted here to calculate $D^* = [d_1^*, d_2^*, \dots, d_n^*]^T$. Since W is full, the embedded matrix-vector multiplication must be improved using the special structure of W . At last, we get the required binary labels straightforwardly by setting the i th label to be *fore* if $d_i^* \geq 0$, or *back* if $d_i^* < 0$. Here is the outline of our new segmentation algorithm:

Algorithm 1 Color image segmentation based on MRF-MAP

- 1: Clustering all colors into m classes; compute $S(c(p), c(q))$ for each pair of adjacent pixels p and q ;
 - 2: Calculate the largest eigenvector $D^* = [d_1^*, d_2^*, \dots, d_n^*]^T$ of the matrix W with *Lanczos* eigensolver, where we obtain the product $R = [r_1, r_2, \dots, r_n]^T$ of W and any vector $V = [v_1, v_2, \dots, v_n]^T$ as
 - let $\varphi = 0$, and $\theta_i = 0$ for all $i = 1, 2, \dots, m$;
 - for $k = 1$ to n , let $\varphi \leftarrow \varphi - \frac{5}{2n} \cdot v_k$, $\theta_{c(k)} \leftarrow \theta_{c(k)} + \frac{5}{2n_{c(k)}} \cdot v_k$, and $\mu_k = \sum_{j:(j,k) \in N} S(j, k) \cdot v_j$;
 - for $k = 1$ to n , $r_k \leftarrow \varphi + \theta_{c(k)} + (\frac{5}{2n} - \frac{5}{2n_{c(k)}}) \cdot v_k$.
 - 3: Output the label of the k th pixel as *back* if $d_k^* < 0$, or *fore* if $d_k^* \geq 0$.
-

The analysis on the computational complexity is rather straightforward: the step 1, 3 can be finished in $O(n)$ time; the matrix-vector multiplication can also be finished in $O(n)$ time, so the *Lanczos* algorithm revoked in step 2 can be finished in $O(n \cdot d)$ time, where d is the amount of performed iterations solving the eigenvector. Therefore the total time complexity of is $O(n \cdot d)$. Since d is irrelevant with n and empirically always less than a certain constant, the new segmentation algorithm is practically a nearly linear one.

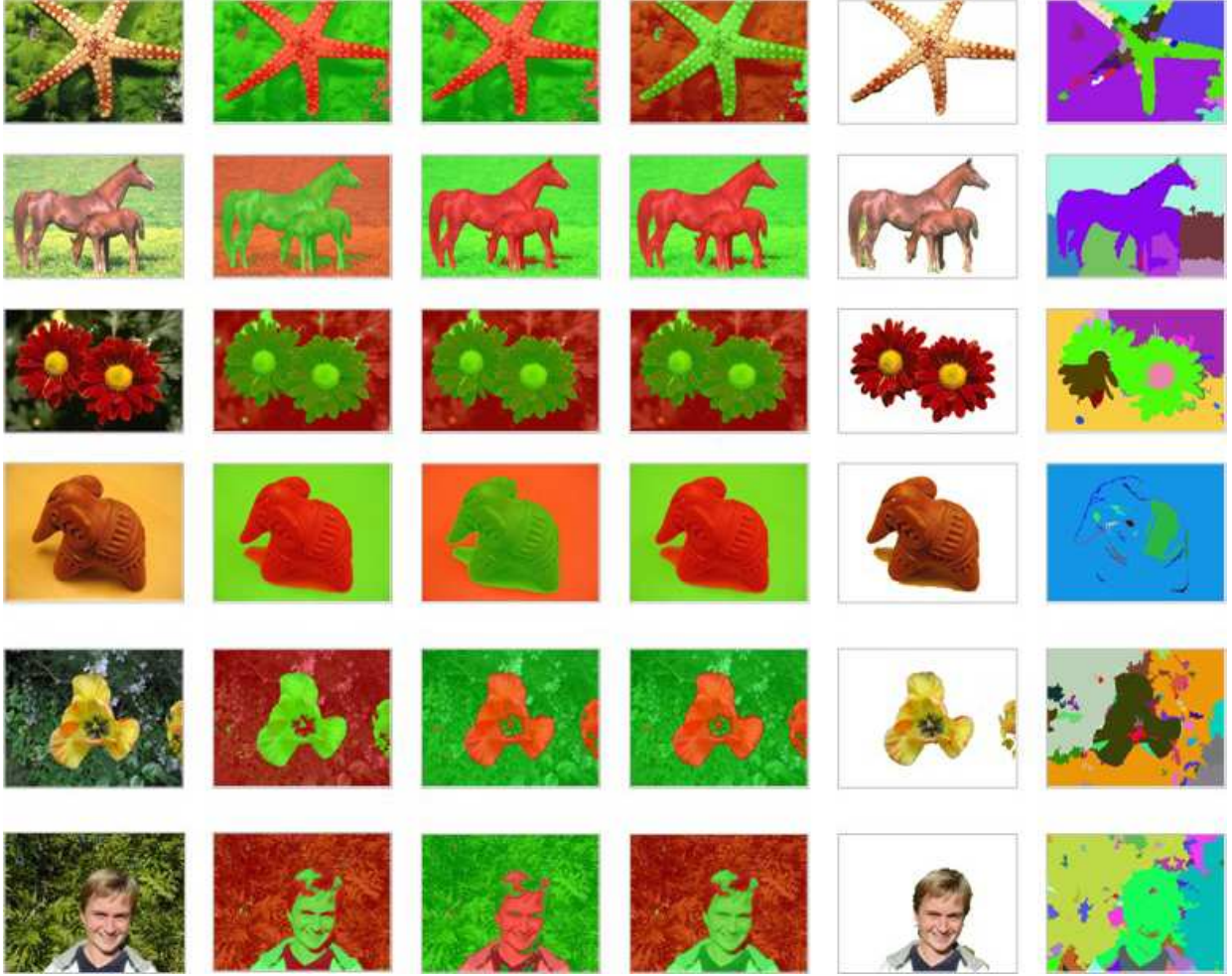


Fig. 2. Comparison of the segmentation results by our method and other two methods. (Leftmost: original images. Other columns from left to right: segmentation results by our method with $\lambda = 1, 5, 10$, by the method in [8] and [14].)

IV. EXPERIMENTS

The experiments are finished on color images chosen from two segmentation datasets from Berkeley and MS research at Cambridge, together with the source codes developed in Matlab. The color clustering method adopted in step 1 is from [13], and all of the amounts of color classes are set to be 16 for these images. λ varies from 1 to 10, and all smoothness terms are simply set to be 1 over all adjacent pixels despite their colors, to say, $S(\cdot, \cdot) \equiv 1$.

These images are intentionally chosen to be of splendid colors and delicate local details, so to verify the performance of our new method facing different challenges. Two state-of-art segmentation algorithms, one supervised [8] and another unsupervised [14], are chosen in the control experiments to examine the

segmentation quality of our new method. In general, all segmentation results of our new method are basically acceptable in the experiments. Especially, the perceptually outstanding objects, if any in the test images, are usually figured out of the underlaying scenes accurately. There are six groups of typical experiment results in Fig. 2: each group includes five different segmentation results of an identical color image, three of them are all from our method but with different λ 's, whereas the rest two are from the other two methods as the comparison. The embedded objects, striking either for fresh colors or large continuous shapes, are precisely outlined by our new method in Fig. 2. Its segmentation quality, is considerably close to the user-interactive-styled method in [8], but much better than the unsupervised one in [14].

Consequentially, there are some noticeable slight differences between the segmentation results on different λ 's of our new method. There are more isolated, but vivid pieces when $\lambda = 1$, and simultaneously the segmentation boundaries are more likely located on the desired edges of the objects. However, it also brings too much emphasis on these discontinuous line segments, and results in much more isolated pieces in the segmentations. When λ varies from 1 to 5, then to 10, it is shown the segmentation boundaries become smoother and smoother on the cost of losing the elaborate details, and the two segmentation zones are more close to each other in sizes when $\lambda = 10$. The reason for that is, the smoothness terms become larger and larger quantitatively so that the continuity of the segmentations is emphasized much more, therefore it is more inclined to cut two pieces of the the same sizes and flat boundaries to reach the minimum energies.

V. CONCLUSIONS

In this letter, a new unsupervised MRF-MAP-based segmentation algorithm is introduced. By introducing an reasonable approximation to the data terms, the energy functions could be minimized remarkably, however, without any supervision. The new method is able to obtain the high-quality segmentation results, as well as the high computational efficiency. The future work includes the investigating the computational hardness of MRF-MAP, and extending the new method to video segmentations and multiple-labeled segmentations.

REFERENCES

- [1] D. L. Pham, C. Xu, and J. L. Prince. Current Methods in Medical Image Segmentation. *Annual Review of Biomedical Engineering*, 2:315-337, 2000.
- [2] N. Alajlan, M. Kamel, and G. Freeman. Geometry-based image retrieval in binary image databases. *IEEE Trans. Pattern Anal. Mach. Intell.*, 30(6):1003-1013, 2008.

- [3] J. Wang, and M. F. Cohen. Image and video matting: A survey. *Foundations and Trends in Computer Graphics and Computer Vision*, 3(2), 2007.
- [4] J. Shotton, J. Winn, C. Rother, and A. Criminisi. Textonboost for image understanding: Multiclass object recognition and segmentation by jointly modeling appearance, shape and context. *International Journal of Computer Vision*, 81(1):2-23, 2009.
- [5] Z. Kato, and T. C. Pong. A Markov random field image segmentation model for color textured images. *Image and vision computing*, 24(3):1103-1114, 2006.
- [6] H. Deng, and D. A. Clausi. Unsupervised image segmentation using a simple MRF model with a new implementation scheme. *Pattern recognition*, 37(4):2323-2335, 2004.
- [7] G. Celeux, F. Forbes, and N. Peyrard. EM procedures using mean field-like approximations for Markov model-based image segmentation. *Pattern recognition*, 36(2):131-144, 2003.
- [8] C. Rother, V. Kolmogorov, and A. Blake. GrabCut: Interactive Foreground Extraction using Iterated Graph Cuts. *ACM transactions on graphics*, 23(3):307-312, 2004.
- [9] A. Blake, C. Rother, M. Brown, P. Perez, and P. Torr. Interactive Image Segmentation Using an Adaptive GMMRF Model. *European Conference on Computer Vision*, 428-441, 2004.
- [10] J. Liu, J. Sun, and H. Shum. Paint Selection. *ACM transactions on graphics*, 28(3):1-7, 2009.
- [11] R. Szeliski, R. Zabih, D. Scharstein, and etc. A comparative study of energy minimization methods for markov random fields with smoothness-based priors. *IEEE trans. on Pattern Analysis and Machine Intelligence*, 30(6):1068-1980, 2008.
- [12] G. Golub, and C. Van Loan. *Matrix Computations*, John Hopkins Press, 1996.
- [13] M. T. Orchard, and A. B. Bouman. Color quantization of images. *IEEE trans. on Signal Processing*, 39(12):2677-2690, 1991.
- [14] P. F. Felzenszwalb, and D. P. Huttenlocher. Efficient Graph-Based Image Segmentation. *International Journal of Computer Vision*, 59(2):167-181, 2004.